Intent Imitation using Wearable Motion Capturing System with On-line Teaching of Task Attention

Tetsunari Inamura, Naoki Kojo, Tomoyuki Sonoda, Kazuyuki Sakamoto, Kei Okada and Masayuki Inaba

Department of Mechano-Informatics
University of Tokyo

Abstract—In order for humanoids to imitate humans behavior, it is important to extract a needful parameter for target of imitation. Especially in daily-life environment, only simple joint angles are insufficiency because position and posture of hands and remarkable point of target object are needed for intent imitation. In this paper, we describe a development methods of motion capturing system with interactive teaching of task attention, and show its feasibility in daily-life environments.

Index Terms—Intent Imitation, Humanoids, Attention of Task, Motion Capture Systems.

I. INTRODUCTION

Recently, imitation skill for humanoids is gaining a great deal of attention because it is said that the imitation function is the most primitive and fundamental factor of intelligence[1]. Satoh et al have started a research project of the robotic imitation, and proposed that the intent imitation function could be a breakthrough for humanoids and artificial intelligence. Fig.1 shows the research map of the project, in which the intent imitation is located as the final goal of the research project.

Intent imitation is a higher conception against simple imitation such as copying of motor command. In the intent imitation, robots have to recognize users’ intent and modify the original motion patterns so as to achieve the purpose with consideration of difference of physical conditions between humans and humanoids. Modeling of users’ intent is important for such imitation, however it is difficult to acquire and describe the intention by only observation. There are several researches on motion generation for humanoids and CG characters using motion capturing system, however, developer must embed the “intent” for the system. Therefore, almost all results have focused on dancing and walking behavior, which do not need consideration of relationship between humanoids’ body and environmental objects.

If motion capturing systems could observe the “intent” of users, humanoids would generate more natural and reasonable behavior for complex tasks in real world. In this paper, we propose an interactive learning mechanism from a viewpoint that interaction between learner and teacher is effective for the acquisition and modification of intent models. For the mechanism, we also propose primitives of attention points, namely primitive intent in daily life behaviors. The interactive learning mechanism enables robots to develop purposive behavior with combination of the taught attention points. We also introduce a wearable motion capturing system for interactive on-line teaching, and a humanoid with the interactive learning mechanism.

II. INTENT IMITATION AND INTERACTIVE TEACHING OF ATTENTION POINT

A. Attention point of daily life tasks

The main target tasks of this research are daily life behavior, such as handling of plate-wares, cleaning of furniture, operation of home information appliances, and so on. In such behavior, intent imitation is needed, because it is difficult for robots to achieve these tasks using only trajectory of hand and joints. Therefore, the robots have to observe not only the trajectories of hand and joints, but also the relationship between humanoid and target objects in order to achieve the tasks with reasonable result. Generally speaking, “skill” is the most important factor for the achievement of tasks, however, we boil down to the question of attention point control.

In this paper, attention point means target factors of imitation, in order words, primitive intent. There are many imitation point for humanoids such as joint trajectories, relationship between self-body and target objects, gaze point of cameras, sensor feedback rules, and so on. Conventional researches of robotic imitation have treated the trajectories and self-behaviors. In contrast, we focus on imitation of other factors, such as handling objects so as to achieve tasks.

Fig. 1. A research map of robotic imitation
We considered several attention points as Table I. These attention points are selected from viewpoint of daily life environment, such as handling of furniture and appliances. Each attention points have condition. 1) Following, 2) Constraint and 3) Disregard are the contents of the conditions.

B. Following condition

a) Following of the relationship between end effector and target objects: Following of the trajectories of target objects and end effector is effective for reaching to the target objects and grasping them with accuracy. It should be appreciated that the unsuitable poses are rejected by kinematic constraint. For example, when inverse kinematics could not be solved, the humanoid keeps a previous pose. Figure 3 shows a situation that a humanoid pick up a kettle with human’s performance.

C. Constraint condition

This constraint is needed for pouring behavior and grasping vertical hand rails, and so on. Users motion, especially gesture motion always differs from real behavior because the gesture motion does not interact with target object, therefore some modification is needed for the original gesture motion. Constraint condition is the most useful modification for such motions. The Constraint condition consists of horizontal, collinear and relative position/posture constraint.

b) Collinear constraint of both hands: Figure 4 shows a situation that there is collinear constraint between both hands. This constraint is used for situations in which humanoids are going to grasp stick with both hands, such as brooms.

c) Constraint of relative position and posture between both hands: Figure 5 shows a situation in which the constraint of relative position and posture between both hands is activated. This constraint condition is needed in which the humanoids are going to hold boxed with both hands.

The relative position and posture constraint is also used when the robot pour liquid matter into some receivers. Figure
### TABLE I
**PRIMITIVE ATTENTION POINTS IN DAILY LIFE BEHAVIOR**

<table>
<thead>
<tr>
<th>Attention Point /<em>(primitive intent)</em>/</th>
<th>Condition</th>
<th>Typical Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position of end effector</td>
<td>Constraint</td>
<td>Pouring water</td>
</tr>
<tr>
<td>Posture of end effector</td>
<td>Following</td>
<td>Grasping a glass with water</td>
</tr>
<tr>
<td>Relative position and posture between both hands</td>
<td>Following</td>
<td>Holding boxes by both hands</td>
</tr>
<tr>
<td>Relative position between hand and target</td>
<td>Following</td>
<td>Press of buttons</td>
</tr>
<tr>
<td>Horizontal constraint of position</td>
<td>Constraint</td>
<td>Polishing tables</td>
</tr>
<tr>
<td>Collinear constraint of both hands</td>
<td>Constraint</td>
<td>Holding sticks with both hands</td>
</tr>
<tr>
<td>Vertical constraint of both hands</td>
<td>Constraint</td>
<td>Wiping windows</td>
</tr>
<tr>
<td>Instruction of ignoring point</td>
<td>Disregard</td>
<td>Removing from attention points</td>
</tr>
</tbody>
</table>

6 shows a situation that the humanoid is going to pour water with a pot.

**d) Horizontal constraint of end effectors:** Figure 7 shows a situation in which the horizontal constraint of end effector. This constraint condition is needed in which humanoids are going to polish and sweep desks with clothes. The humanoid in Fig.7 is also going to keep the posture of end effector in order to fit one hand to the surface of the desk.

**e) Disregard:** The Disregard condition is used for a situation in which the user want to teach single-handed task. In such a situation motion patterns of another unused hand are ignored.

### D. On-line and interactive intent imitation system

Figure 8 shows the whole system of the interactive intent imitation system. Solid line indicates flow of motion patterns, broken line indicates flow of task attention information. An user (teacher) instructs example motions in real-time, with giving voice commands for task attention. The details of capture system is shown in Section 3. The motion patterns performed by the teacher are sent to a motion modifier. The motion modifier accepts task attention from voice recognizer. Basically, the motion patterns are modified with consideration of kinematic conditions of the human and the humanoid. The details of the motion modifier is described in Section IV. The motion patterns and task attention information are also sent to learning/recognition subsystem. The subsystem segments the motion patterns with the help of task attention information. Segmented motion patterns are learn with label of the task attention. After the learning, the recognition subsystem can generate suitable motion patterns even if the teacher performs partial and uncertain motions. The generated motion patterns are also used for control of the humanoid. The learning/recognition subsystem is explained in Section V.

### III. WEARABLE MOTION CAPTURING SYSTEM WITH ON-LINE TEACHING OF ATTENTION POINT

Recently, motion capturing systems are widely used in behavior learning and teaching for humanoids. Almost all motion capturing systems adopt optical device or magnetic device, however, they are inconvenient in daily life environments because of the restriction of movable area. In this paper, we adopted a wearable motion capturing system without such a restrictions. On-line interactive teaching function is also added on the motion capturing system for the teaching of attention points.

**A. Wearable motion capturing system**

The used motion capture system is “GypsyGyro” manufactured by Spice Inc. and Animazoo Inc. This capture device use 18 gyro sensors and these sensors are attached on testee body as shown in Fig.9. Each gyro sensor can measure acceleration.
for three degrees, and send the measured data to center unit via wireless transmission. Sampling rate of the measurement is 120[fps], resolution is 0.03[deg] and the maximum measure error is about 1[deg]. These properties satisfy the use in daily life environment and imitation of the objective behavior. In other words, there is no need to measure accurate posture, because the conditions and attention is the most important information for the humanoid.

The wearable motion capture system enables humanoids to imitate users’ behavior in anywhere without any restrictions of movable area. We have confirmed the experiment in outdoor environment as shown in Fig.2. With the help of the system, wide range of daily life behaviors can become to be target of the robotic imitation.

B. Attention teaching with voice recognition

When human tell attention points for the humanoid during using motion capturing system, voice command is the most suitable way to communicate with the humanoid. We adopted free software package “Julius/Julian”[2] as voice recognition subsystem. The “Julius/Julian” can accept grammar model for upgrading of recognition rate. Some grammar and sentence correspond with the conditions in Table.I are registered on the voice recognition system. User can instruct various conditions and attention points using the voice command.

IV. ON-LINE MOTION MODIFICATION BASED ON TASK ATTENTION AND ENVIRONMENT MODEL

A. On-line modification of motion patterns

The humanoids have to modify the original motion patterns in order to satisfy the condition of attention points. The modification of the motion patterns have to consider handling of target objects, self-body collision and consistency of the purpose of task. We have developed motion generation system in order for the humanoids to act naturally in daily life environments[3]. The system can modify the original motion patterns so as not to break the consistencies.

Figure 10 shows the modification strategy. Joint angles of the performer measured by the motion capture system is sent to the kinematic calculation module. In the module, positions and postures of focused hands is used for the motion modification. With the task attention information, original positions and postures of hands are modified. Final motion patterns of the humanoid is generated with forward kinematics with the modified positions and postures of hands.

V. SYMBOLIZATION OF MULTI-SENSORY DATA AND INTENT IMITATION

So far, we have proposed a mathematical model that abstracts the whole body motions as symbols, generates motion patterns from the symbols, and distinguishes motion patterns based on the symbols. In other words, it is a functional
realization of the mirror neurons and the mimesis theory. For the integration of abstract, recognition and generation, the hidden Markov model (HMM) is used. One as observer would view a motion pattern of the other as the performer, the observer acquires a symbol of the motion pattern. He recognizes similar motion patterns and even generates it by himself. One HMM is assigned for a kind of behavior. We call the HMM as symbol representation[4].

Another characteristics of the symbol representation is that geometric symbol space can be constructed which contains relative distance information among symbols. In order words, meaning and tendency of behaviors are described as geometric relationship of the space constitution[5]. The humanoid can recognize unknown behavior as a point in geometric space, therefore distances between the point of unknown behavior and points of known behaviors indicate the status of recognition. The configuration of symbolization system is shown in Fig.11.

VI. EXPERIMENT OF INTENT IMITATION ON A HUMANOID ROBOT: HRP2W

We adopted HRP-2W[6] as a humanoid robot platform for the interactive motion acquisition and objective behavior imitation. One of the concepts of the platform is that the researcher can focus on the intelligence layer without consideration of delicate balance control.

This kind of humanoids with wheel unit have already proposed [7][8]. The differences between those research are continuous act for the storage of shared experiences, and multiple sensors for the plentiful experiences. The following lists are the loaded sensors on the humanoid platform;

- 20DOFs: 3 for each shoulder, 1 for each elbow, 3 for each wrist, 1 for finger on each hand, 2 for head, 2 for waist.
- Binocular color cameras for stereo vision.
- Stereo microphones for speech dialogue and sound source orientation.
- A speaker for speech utterance.
- Force sensor for six axes on both hands.
- Independence system based on batteries with large capacity and wireless LAN.

Fig. 11. Recognition and Learning of multi-sensory data using proto-symbol space

Fig. 12. A Humanoid Platform: HRP-2W

A. On-line imitation experiments

We have practiced teaching and generation of daily life behaviors to confirm effectiveness of the proposed method. In the teaching phase, pouring water into a glass, throwing a ball and swinging both hands are selected. For the pouring behavior, the robot uses the restriction condition of relative restriction condition of horizontal constraint.

Figure 13 and 14 show the result of on-line modification of performed motions. In Figure 13, the user instruct to the robot that the attention point of “constraint of relative position and posture” should be used. Then, the original performed motion is modified not to spill water. With the help of the motion modifier, if the user performs unsuitable motion as shown middle picture in Fig.13, the robot success to pour water into a glass.

In Figure 14, the user instruct to the robot that the attention point of “constraint of horizontal position” should be used. Then, the original performed motion is modified to keep a certain height.

B. Behavior acquisition and recalling Experiments

Next, we confirmed the learning and recalling subsystem. In the learning phase, observed joint angles for 20 joints are used for the HMM based symbolization subsystem. Time series of the joint angles are abstracted as static points in the geometric symbol space.

For the recognition, the humanoid always calculate similarity between present performed behavior and learn behaviors. The similarity is calculated as distances between the state points using the geometric symbol space. A state point which
is located at the minimum distance from a state point of the performed motion, is selected as the most suitable behavior for current situation.

The humanoid can recognize which behavior should be selected using sensor information in the shortest time. After the recognition, original motion patterns could be generated. As well as the on-line motion modification, the recalled motion patterns are modified with attention points which is the result of the recognition process.

Figure 15 shows the time change of distance between known proto-symbol and observed behavior. Each line indicates pouring, carrying, wiping and putting respectively. An example behavior is as follows; (1) Pouring water into a glass, (2) Carrying a glass without spilling, (3) Wiping a desk with a rag, (4) Put the glass on the desk.

VII. CONCLUSIONS

In this paper, we focused on a decision of attention points in order for humanoid robots to imitate humans’ objective behavior in daily life environment. For the purpose, we developed a wearable motion capturing system with interactive teaching function of attention points, which enables users to instruct motion patterns and the important point of the behavior for achievement of the task.

In current stage, taught attention points are just stored in memory, and they are just referred in behavior generation phase. The modification of original rough motion patterns into reasonable motion patterns which satisfies the aim of behavior, shows a convenient performance, however, it is desirable that the humanoid learns which attention points is the most suitable condition depending on the situation. We are now planning to adopt the learning framework mentioned in Section V for the problem. HMM based behavior symbolization system can treat several kinds of modalities such as vision, force, joint and distance sensors. Therefore, if the selection of attention points can be described as sensor information, the strategy of attention selection is learn by humanoids without any modification of the system.

Such a integration enables the system to be applied to the learning and teaching in more natural way. For example, if humanoids can recognize the constraint condition, following condition and so on, users get free of instruction of the attention points. And such a situation can be regarded as a huge step forward to the realization of objective imitation for humanoid robots.

REFERENCES